SYNOPSIS

Data mining is an important branch of Artificial intelligence (AI) that searches for valuable information in a large database commonly known as knowledge discovery. Due to significant strides in information processing technology there exists several data mining techniques like clustering, classification, prediction, association, characterization, rule generation, and pattern recognition, etc. for a variety of databases that include object oriented, relational, temporal, transactional types. The object of study in the present research is the time series database which is an interesting and important area of research as it belongs to many real world problems occurring in finance, industry, energy markets, medical diagnostics, geotechnical engineering, etc. It is to be noted that the time series data bases exhibit time varying characteristics and their statistical characteristics are also nonstationary. A time series database comprises a large number of data points representing. a certain quantitative value of the data measured at a particular instant of time. The analysis of such data is an important task to discover pertinent knowledge in data and hence time series data mining assumes significance and usefulness. Some of the time series databases belonging to the domains of bioinformatics, financial, and energy markets require accurate prediction and decision making. Earlier techniques for mining of large time series data bases usually involved statistical methods that fitted the observed data to a simple mathematical model and subsequently used the fitted model for future prediction and pattern discovery. However, due to the chaotic nature of the time series data, the statistical methods failed to produce the required accuracy thus paving the way for applications of intelligent systems and machine learning techniques for data mining and knowledge discovery in time series databases. The present research in the area of mining of time series data will consider the two most important aspects like prediction and classification which will also include clustering. In particular two important time series databases belonging to financial and energy markets are considered for advanced studies for developing new techniques for forecasting and classification. Financial time series data being quantitative and nonstationary in nature poses a serious challenge for its prediction. Also the presence of noise in the data produces more difficulties for achieving the necessary accuracy in short-term and long-term prediction. Also in energy markets undergoing deregulation and

restructuring, electricity price exhibits complex behavior which is nonstationary, nonlinear and time varying and therefore requires a more accurate and robust prediction process.

Although numerous studies in time series data bases like the financial time series have been undertaken there still remain several challenges to accurately predict its future behavior based on its past performance. Price index forecasting is one of the most important problems in financial markets that make the investors worry about their investments. Stock market prediction is of great interest to stock investors and traders due to high profit in trading the stocks. Thus the prediction of stock market indices and its analysis are important to ascertain whether the next day's closing price would increase or decrease. In the past decades, the stock market prediction has played a vital role for the investment brokers and individual investors and researchers are on the constant lookout for a reliable method for predicting stock market trends. Traditionally statistical approaches like ARMA, ARIMA, GARCH models have been used by researchers to predict one day ahead stock market indices, but they are found to produce unsatisfactory performance and this is primarily due to the chaotic nature of the stock market variations. In their search for methods for addressing these shortcomings, many market analysts, traders, and researchers have investigated the various intelligent system techniques for analyzing the stock markets and making trading decisions. The various tools of computational intelligence include Artificial neural networks (ANN) , Fuzzy Logic Systems, Support Vector Machines (SVM), etc. The different types of ANNs used for stock market analysis include Radial Basis Function neural network (RBFNN), Recurrent neural network (RNN), Multilayer Perceptron network (MLP), Generalized Regression neural network (GRNN), Random Vector Functional Link Net (FLANN), Local Linear Wavelet neural network (LLWNN), Wavelet neural network (WNN), etc.. These neural models, however, do not perform so successfully due to the dimensionality, volatility, and noise problems associated with the stock price data.

As the primary objective of the thesis is to develop a robust neural information system to efficiently mine two important data bases like the stock market indices, electricity prices and its quality, the proposed research presents studies on a suitable neural architecture and its learning paradigm. In this regard recurrent neural network architecture is examined for time series prediction, since these types of networks can effectively

represent a time varying dynamical system like the stock and energy markets. A recurrent neural network uses local or global feedback producing a recursive nature unlike the other types of neural networks and these results in a smaller architecture. It is well known that a dynamic neuron can be realized using an IIR filter in its output path and a number of such neurons constitute the dynamic or recurrent neural network (DNN). On the other hand globally recurrent architectures use feedback from the output neurons and have been also used for system identification and prediction. In comparison, the multilayered feed-forward neural networks exhibit static nonlinear architectures and may not be suitable to capture the nonlinear dynamic behavior as required in chaotic time series forecasting. Further the multilayered feed-forward neural networks are usually trained by gradient descent and back-propagation algorithms which are known to suffer from slow convergence, local minima, and over fitting problems. Contrary to these the recurrent neural networks are trained by real-time recurrent learning algorithm (RTRL) [69] and back-propagation through time (BPTT), etc. but they suffer from slow convergence and may not be able to provide the required accuracy. For a nonlinear system like the DNN, the weight estimation problem can be tackled using the well known Kalman filters like the extended Kalman filter (EKF) and the Unscented Kalman filter (UKF). Among the two filters the UKF produces more accurate prediction and does not require a Jacobean matrix to be inverted at every iterations.

However, estimation accuracy of both the EKF and UKF algorithms depend on the correct choice of noise covariance matrices and the estimation can get degraded when there are sudden changes in the input data to the neural network. Thus in this research evolutionary optimization techniques like the particle swarm optimization (PSO) or differential evolution (DE) are considered to be embedded into the UKF algorithm in a hybrid manner to produce a more robust forecasting system. Alternatively the evolutionary algorithms like PSO and DE ,etc. can be directly applied to train the weights of the dynamic neural network. But the time required to achieve optimized weight parameters is much higher than the EKF or UKF based learning algorithms. Another simpler learning approach is to use the sliding mode strategy for forecasting of both financial and energy time series databases. It is well known from control system design that the sliding mode approach guarantees robust stability, fast convergence in the presence of uncertainties thereby

providing a very efficient learning strategy for the neural architectures. In the sliding mode control (SMC) strategy the motion of a system is constrained to a hyperplane known as the sliding surface that reduces the learning error to zero. Also its stability can be ascertained using Lyapunov's direct stability theorem.

After studying the various learning paradigms for the dynamic neural network the stock price indices for a number of stocks like the BSE Sensex, Nifty, Oracle, and IBM, etc. are considered for a day ahead prediction to a month ahead prediction. Further other attributes like stock trend and volatility which are also of immense importance for the investors to take decisions on trading of the stocks have been considered for detailed studies. For volatility forecasting the well known GARCH model is hybridized with a functional link neural network and simple IIR based neural architecture is used to train the weight parameters. The applications of DNN to other time series databases like the foreign exchange and electricity price prediction giving new and interesting results are also considered in this research. During the last several years the power industry in many advanced countries like U.K. and U.S.A. has introduced restructuring and deregulation and therefore the electrical energy is treated as a commodity and forecasting the price of electricity on a short-term basis has assumed great importance. Electricity price is the most important information in electric power market, especially locational marginal price (LMP). Forecasting electricity parameters such as energy and power have become major issues in power markets.

Apart from the prediction of electricity price fluctuations, another important aspect of energy market studies is to determine whether the electricity supplied to the customers is of right quality or not. The overall price of electricity supplied to customers under deregulated and restructured energy market depends not only on the electricity price variations but also qualitative nature of electricity. In electricity distribution networks the voltage and current signals are seldom steady and their root mean squares (R.M.S.) values fluctuate randomly along with their waveforms primarily due to the wide use of power electronic devices by the end-users. The power quality disturbance signals constitute a time series database and exhibits a fluctuating pattern that requires a serious study about its behavioral pattern and recognition. Also its analysis encompasses several technologies like digital signal processing, VLSI, power electronics and embedded systems, software

technology, and the data mining. Practical devices involving both software and hardware concepts for monitoring electricity supply quality are making inroads to the market place particularly to the industrial concerns. Also hybrid techniques involving both the advanced signal processing like the Wavelet and S-Transforms and machine intelligence are required to classify both the single and simultaneous electricity supply disturbance patterns and bring out certain similarities in the disturbance time series database. In this research a modified generalized S-Transform has been used to extract pertinent features from the power signal time series data and a PSO based K-means clustering analysis and a Decision tree are used to identify the disturbance patterns in the data. It is to be noted here that the time series clustering has been very effective in providing pertinent and useful information in various application domains as a part of temporal data mining research.

The main objectives of the thesis and major contributions are summarized as:

1. To develop new dynamic neural network architecture along with an adaptive filter based evolutionary learning algorithm for short-term and long-term prediction of stock market indices. The purpose of this study is to improve further prediction accuracy using several performance metrics.
2. To develop new rules for the prediction of stock market trends using a selected number of technical indicators, which will be ultimately useful for decision making strategies to buy or sell stocks.
3. To develop a new hybrid DNN and functional block based architecture for volatility prediction stock market closing prices with inputs from a GARCH model.
4. New forecasting models for foreign exchange and electricity price time series prediction are developed .
5. A new evolutionary clustering and classification algorithm is developed based on a modified S-Transform and decision tree techniques.

Finally, the thesis is organized in six chapters and apart from the introduction in Chapter-1, other chapters describe the various aspects of mining financial and electricity price and quality time series databases using evolutionary neural information systems. Concluding remarks given in the final chapter provide comprehensive summary of the proposed intelligent system techniques and new learning paradigms applied for the prediction of

stock market closing price indices, volatility and identifying market trends. Also it outlines some of the new data mining research areas in the financial and energy markets to be explored further.

(Ranjeeta Bisoi) **(Prof.(Dr.) P.K. Dash**

Candidate **Supervisor**